REMARKS

Claims 1, 3-7, 9-12 have been amended to further define the claimed invention. Claims 2 and 8 have been cancelled.

The specification has been amended to cancel the text in paragraphs 0072 and 0074.

Claims 1-12 have been rejected under 35 U.S.C. 102(e) as being anticipated by Shlomot (US patent 6,377,931).

First, it is respectfully submitted that the amended independent claims are defined over the prior art.

In particular, claim 1 recites an audio decoding device including a jitter buffer comprising a plurality of buffer portions for storing a received packet, and decoding means for decoding the packet stored in the jitter buffer, wherein the received packet is stored in a position corresponding to its packet number in the jitter buffer by using a packet number of a packet stored in a buffer portion at an output end of the jitter buffer as a reference packet number,

the audio decoding device comprising:

playback speed change means for changing, with respect to a decoded audio signal obtained by the decoding means, the playback speed thereof;

an output buffer for temporarily storing a digital audio signal outputted from the playback speed change means;

means for reading out the digital audio signals stored in the output buffer at predetermined time intervals;

playback speed control means for controlling the playback speed change means on the basis of the position in which the received buffer is stored in the jitter buffer; and

decoding timing control means for controlling the timing of decoding by the decoding means on the basis of the amount of data stored in the output buffer;

wherein

a first region, a second region and a third region are set within the jitter buffer, the first region being composed of a given number of buffer portions from the output end of the jitter buffer, the second region being composed of a given number of buffer portions and lying between the first region and an opposite end of the output end in the jitter buffer, and the third region being composed of a given number of buffer portions and lying between the second region and the opposite end of the output end in the jitter buffer; and

the playback speed control means controls the playback speed change means such that the playback speed is reduced when the received packet is stored in the first region of the jitter buffer, while controlling the playback speed change means such that the playback speed is increased when the received packets are stored in the third region of the jitter buffer a predetermined consecutive number of times or more.

Independent claim 4 recites an audio decoding device including a jitter buffer comprising a plurality of buffer portions for storing a received packet, and decoding means for decoding the packet stored in the jitter buffer, wherein

the received packet is stored in a position corresponding to its packet number in the jitter buffer by using a packet number of a packet stored in a buffer portion at an output end of the jitter buffer as a reference packet number, and

a first region, a second region and a third region are set within the jitter buffer, the first region being composed of a given number of buffer portions from the output end of the jitter buffer, the second region being composed of a given number of buffer portions and lying between the first region and an opposite end of the output end in the jitter buffer, and the third region being composed of a given number of buffer portions and lying between the second region and the opposite end of the output end in the jitter buffer,

the audio decoding device comprising:

delay time control means for carrying out such control that a delay time period elapsed from the time when the packet is stored in the jitter buffer until the packet is decoded is lengthened when the received packet is stored in the first region of the jitter buffer, while carrying out such control that a delay time period elapsed from the time when the packet is stored in the jitter buffer until the packet is decoded is shortened when the received packets are stored in the third region of the jitter buffer a predetermined consecutive number of times or more.

Independent claim 7 recites a network telephone set including a jitter buffer comprising a plurality of buffer portions for storing a received packet, and decoding means for decoding the packet stored in the jitter buffer, wherein the received packet is stored in a position corresponding to its packet number in the jitter buffer by using a packet number of a packet stored in a buffer portion at an output end of the jitter buffer as a reference packet number,

the network telephone set comprising:

playback speed change means for changing, with respect to a decoded audio signal obtained by the decoding means, the playback speed thereof;

an output buffer for temporarily storing a digital audio signal outputted from the playback speed change means;

means for reading out the digital audio signals stored in the output buffer at predetermined time intervals;

playback speed control means for controlling the playback speed change means on the basis of the position in which the received packet is stored in the jitter buffer; and

decoding timing control means for controlling the timing of decoding by the decoding means on the basis of the amount of data stored in the output buffer;

wherein

a first region, a second region and a third region are set within the jitter buffer, the first region being composed of a given number of buffer portions from the output end of the jitter buffer, the second region being composed of a given number of buffer portions and lying between the first region and an opposite end of the output end in the jitter buffer, and the third region being composed of a given number of buffer portions and lying between the second region and the opposite end of the output end in the jitter buffer; and

the playback speed control means controls the playback speed change means such that the playback speed is reduced when the received packet is stored in the first region of the jitter

buffer, while controlling the playback speed change means such that the playback speed is increased when the received packets are stored in the third region of the jitter buffer a predetermined consecutive number of times or more.

Independent claim 10 recites a network telephone set including a jitter buffer comprising a plurality of buffer portions for storing a received packet, and decoding means for decoding the packet stored in the jitter buffer, wherein the received packet is stored in a position corresponding to its packet number in the jitter buffer by using a packet number of a packet stored in a buffer portion at an output end of the jitter buffer as a reference packet number,

the network telephone set comprising:

delay time control means for carrying out such control that a delay time period elapsed from the time when the packet is stored in the jitter buffer until the packet is decoded is lengthened when the received packet is stored in the first region of the jitter buffer, while carrying out such control that a delay time period elapsed from the time when the packet is stored in the jitter buffer until the packet is decoded is shortened when the received packets are stored in the third region of the jitter buffer a predetermined consecutive number of times or more.

It is respectfully submitted that the prior art of record does not disclose the claimed arrangement.

Considering the reference, Shlomot discloses compressing a plurality of audio packets to accelerate the playback of the plurality of audio packets when a rate of receipt of audio packets is greater than a predetermined upper replay rate, and decompressing the plurality of audio packets

to decelerate the playback of the plurality of audio packets when the rate of receipt of the plurality of audio packets is less than a predetermined lower replay rate (see claims 1-6 of Shlomot, and col. 5, line 44-col. 6, line 18).

Further, Shlomot discloses compressing a plurality of .audio packets to accelerate the playback of the plurality of audio packets when the jitter buffer 260 detects an overflow danger, and decompressing the plurality of audio packets to decelerate the playback of the plurality of audio packets when the jitter buffer 260 detects an underflow danger (see claim 11 of Shlomot, and col. 5, line 44-col.7, line 20).

Specifically, the playback speed is controlled based on the number of audio packets stored in the jitter buffer 260.

Detecting the overflow danger and the underflow danger is based on a position of the pointer 340. Note that the pointer 340 points to a CSP (packet) to be decoded and played next.

With respect to the jitter buffer 260 of Shlomot (see Fig. 3)

A newly received CSP (packet) is pushed into the jitter buffer 260 from the right. All of the unplayed CSPs in the jitter buffer 260 are shifted one location to the left, and the pointer 340 is also moved one location to the left.

The pointer 340 is shifted one location to the right when one CSP has been decoded and played. When the pointer 340 approaches the F location as shown in Fig. 3, the overflow danger is detected, and when the pointer 340 approaches the S location as shown in Fig. 3, the underflow danger is detected.

By contrast with the prior art, the jitter buffer used in the present invention is configured, for example, as shown in Fig. 1, where the received packets are stored in order of their packet numbers (the numbers in accordance with time series, called sequence numbers) from the left in the buffer portions of the jitter buffer. Specifically, incoming packets from the network are stored in the proper positions (P2-Pn) by referring to the packet number of the packet stored in P1 (the leftmost buffer portion) as a reference packet number. For instance, if the packet number of the packet stored in P1 is "N" and the packet number of the packet which has just arrived is “N+5”, then the packet which has just arrived is stored in P6. After the packet in P1 is taken out and fed to the decoder, the data within the jitter buffer are shifted respectively from P2 to P1, from P3 to P2, from P4 to P3, from P5 to P4 ... Any area without a packet due to change in the arriving order of packets, packet losses and so forth, is retained empty.

In short, the received packets are stored in positions corresponding to their respective packet numbers in the jitter buffer, by using the packet number of the packet stored in the buffer portion at the output end of the jitter buffer as a reference packet number.

The specification describes that "[t]he packets which have arrived are stored in the order of their packet numbers from the left in the buffer portions in the jitter buffer 33". Hence, the received packets are stored in positions corresponding to their respective packet numbers in the jitter buffer, by using the packet number of the packet stored in the buffer portion at the output end of the jitter buffer as a reference packet number.

For example, Figs. 2a-2e (or Figs. 5a-5d) illustrate the distribution of the times when the packets arrive. The distribution curves as shown in Figs. 2a-2e (or Figs. 5a-5d) can be obtained because the received packets are stored in positions corresponding to their respective packet

numbers in the jitter buffer, by using the packet number of the packet stored in the buffer portion at the output end of the jitter buffer as a reference packet number.

For example, paragraph [0007] on page 3 of the specification describes that "[w]hen the fixed delay in the IP network is increased during telephone conversations, the distribution of the packets which arrive at the jitter buffer 101 is moved from S0 to S2, as shown in Fig. 2c. In this case, the packet which arrives at a portion departing from the jitter buffer 101 cannot be outputted to the decoder 102, so that the audio quality is degraded, similarly to the packet loss". In addition, paragraph [0008] on page 3 describes that "[w]hen the amount of jitter in the IP network is increased during telephone conversations, the distribution of the packets which arrive at the jitter buffer 101 is changed from S0 to S3, as shown in Fig. 2d. In this case, the packet which arrives at the portion departing from the jitter buffer 101 cannot be outputted to the decoder 102, so that the audio quality is degraded, similarly to the packet loss".

Incoming packets may be deviated from the jitter buffer 101 in this way because the received packets are stored in positions corresponding to their respective packet numbers in the jitter buffer, by using the packet number of the packet stored in the buffer portion at the output end of the jitter buffer as a reference packet number.

<u>In such a jitter buffer, there is no relation between the position at which an incoming packet is stored and the amount of stored packets. For example. even if one or more packets are stored in the region C of Fig. 8. an incoming packet may be stored in the region A or B. Likewise, even if the region A or B has empty areas, an incoming packet may be stored in the region C.</u>

In the meantime, in the VoIP technique, audio telephone conversation is conducted over the IP network. Therefore, fluctuation (jitter) takes place in the received packets at each terminal. Since jitter greatly affects the quality in audio telephone conversations, a jitter buffer is generally provided to absorb jitter. Concern here is a target amount of data to be stored in the jitter buffer.

If the target data amount to be stored in the jitter buffer is small, jitter cannot be absorbed sufficiently when a large amount of jitter is occurring. This causes breaks in audio due to packet losses, leading to degraded audio quality.

In reverse, when the target amount of data to be stored in the jitter buffer is large, jitter can be absorbed sufficiently even if a large amount of jitter is occurring. However, when a small amount of jitter is occurring, the fixed delay (a time period until voiced speech at the caller terminal is reproduced at the recipient terminal) becomes large because the buffering amount (the number of packets stored in the jitter buffer) is larger than required, preventing smooth conversations. This situation resembles a phenomenon in satellite communications, where it takes some time to receive speaker's voice, preventing smooth conversations.

For these reasons, it is preferable to determine the target amount of data to be stored in the jitter buffer as follows:

(a) Determine the minimum data amount sufficient to absorb actually occurring jitter as a target value.

(b) Determine the target data amount dynamically because jitter varies moment by moment.

When the target number of packets stored in the jitter buffer is fixed as in Shlomot invention, the above feature (a) cannot be achieved. If the network quality is assured and an amount of jitter to be expected to occur is known in advance, fixing the target number of packets to be stored will produce effect. However, if an expected amount of jitter is not known in advance, appropriate operation will not be performed.

In the VoIP technique, packets are transmitted from the terminal to the network at. regular intervals. Packets received from the network are not at regular intervals because of the influence of loads on the network. Figs. 2a-2e illustrates how to absorb this jitter in the jitter buffer.

Figs. 2a-2e of the present application show the changes in the packet arrival conditions due to variations in the loads on the network. When the packet arrival conditions change as shown in Figs. 2b-2e, a buffering amount of the jitter buffer (the number of packets to be stored in the jitter buffer) must be adjusted in the manner as shown in Figs. 5a-5d. Such adjustment is not performed well, by merely setting the target buffering amount of the jitter buffer with a threshold.

In Figs. 5a-5d, the buffering amounts to be stored in the jitter buffer (the number of packets to be stored in the jitter buffer) are different. In order to absorb jitter in the distribution S4 as shown in Fig. 5d, a small buffering amount is sufficient. However, to absorb jitter in the distribution S3 as shown in Fig. 5c, a larger buffering amount is required. Therefore, the target buffering amount should be determined in consideration of the size of jitter actually occurring.

One method of setting the target buffering amount in consideration of the size of jitter actually occurring is to calculate a delay deviation in arrival time of the received packets.

However, the delay deviation cannot be calculated unless a considerable number of packets is received, and the reliability of the calculated delay deviation is decreased while a buffering amount is varying.

The present invention achieves the features of above items (a) and (b) without calculating a delay deviation in arrival time of the received packets.

Specifically, as shown in Figs. 5a-5d, for example, the buffering amount is controlled such that the end of the distribution curve of jitter actually occurring is located at the left end of the jitter buffer (the data outputting side to the decoder).

For such control, the region in the jitter buffer is divided into three regions - A, B and C as shown, for example, in Fig. 8, and it is monitored how the packets are stored in each area to determine dynamically the target buffering amount.

When a received packet is stored in the region A (that may correspond, for example, to the claimed first region), it is considered that the distribution of jitter is shifted to the left compared with its target position, so that control is made to increase the buffering amount. When received packets are stored consecutively in the region C (that may correspond, for example, to the claimed third region) a predetermined number of times or more, it is assumed that the distribution of jitter is shifted to the right compared with its target position. Therefore, control is made to decrease the buffering amount.

<u>Such control makes it possible to determine dynamically an optimum buffering amount by only monitoring a position at which a received packet is stored in the jitter buffer, without</u>

<u>calculating the delay deviation in arrival time of the received packets and without setting a target value of the buffering amount.</u>

In particular, the playback speed control means, as recited in the claims 1 and 7, controls the playback speed change means such that the playback speed is reduced when the received packet is stored in the first region (such as region A of Fig. 8) of the jitter buffer, and the playback speed is increased when the received packets are stored in the third region (such as region C of Fig. 8) of the jitter buffer a predetermined consecutive number of times or more.

In short, the claimed playback speed control means does not control the playback speed by comparing the number of packets stored in the jitter buffer with a threshold, but does so <u>on the basis of a position at which the received packet is stored</u>.

In contrast, Shlomot controls the playback speed based on the number of CSPs stored in the jitter buffer 260 or the rate of receipt of audio packets.

Further, the delay time control means, as recited in the claims 4 and 10, does not control the delay time period elapsed from the time when a packet is stored in the jitter buffer until the packet is decoded, by comparing the number of packets stored in the jitter buffer with a threshold, but does so <u>on the basis of a position at which the received packet is stored</u>.

Specifically, the delay time control means carries out such control that a delay time period elapsed from the time when the packet is stored in the jitter buffer until the packet is decoded is lengthened when the received packet is stored in the first region (such as region A of Fig. 8) of the jitter buffer, while carrying out such control that a delay time period elapsed from the time when the packet is stored in the jitter buffer until the packet is decoded is shortened

when the received packets are stored in the third region (such as region C of Fig. 8) of the jitter buffer a predetermined consecutive number of times or more.

In contrast, Shlomot controls the playback speed based on the number of CSPs stored in the jitter buffer 260 or the rate of receipt of audio packets.

Anticipation, under 35 U.S.C. § 102, requires that each element of a claim in issue be found, either expressly described or under principles of inherency, in a single prior art reference. ***Kalman v. Kimberly-Clark Corp.,*** 713 F.2d 760, 218 USPQ 781 (Fed. Cir. 1983); ***Richardson v. Suzuki Motor Co.,*** 868 F.2d 1226, 9 USPQ2d 1920 (Fed. Cir. 1989) *cert. denied,* 110 S.Ct. 154 (1989). The term "anticipation," in the sense of 35 U.S.C. 102, has acquired the accepted definition meaning "the disclosure in the prior art of a thing substantially identical with the claimed invention." ***In re Schaumann***, 572 F.2d 312, 197 USPQ 5 (CCPA 1978).

As demonstrated above, Shlomot does not expressly or inherently disclose the elements recited in the amended independent claims 1, 4, 7 and 10. The other references of record also does not disclose these arrangements.

Also, the prior art does not disclose the subject matter of the dependent claims 3, 5, 6, 9, 11 and 12.

Hence, claims 1, 3-7, and 9-12 are clearly defined over the prior art.

In view of the foregoing, and in summary, claims 1, 3-7, and 9-12 are considered to be in condition for allowance. Favorable reconsideration of this application, as amended, is respectfully requested.

To the extent necessary, a petition for an extension of time under 37 C.F.R. 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account 500417 and please credit any excess fees to such deposit account.

Respectfully submitted,

McDERMOTT WILL & EMERY LLP

Alexander V. Yampolsky
Registration No. 36,324

600 13th Street, N.W.
Washington, DC 20005-3096
Phone: 202.756.8000 AVY:MWE
Facsimile: 202.756.8087
Date: December 27, 2007

Please recognize our Customer No. 20277 as our correspondence address.